Chapter 4: 4.3, 4.7, 4.13, 4.18, 4.22

Jeremy Ling & Emmanuel Mejia

March 29, 2018

4.3

A chemist wishes to test the effect of four chemical agents on the strength of a particular type of cloth. Because there might be variability from one bolt to another, the chemist decides to use a randomized block design, with the bolts of cloth considered as blocks. She selects five bolts and applies all four chemicals in random order to each bolt. The resulting tensile strengths follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw appropriate conclusions.

```
library(readxl)
cloth <- read_excel("cloth.xlsx")




clothLong=reshape(cloth,
                    varying = c("Bolt1", "Bolt2", "Bolt3",
                                "Bolt4", "Bolt5"),
                    v.names = "Yield", timevar = "Bolt",
                    direction = "long")
```

```
cloth.aov=aov(Yield ~ factor(Bolt) + factor(Chemical), data = clothLong)
summary(cloth.aov)
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)    
## factor(Bolt)      4 157.00   39.25  21.606 2.06e-05 ***
## factor(Chemical)  3  12.95    4.32   2.376    0.121    
## Residuals        12  21.80    1.82                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
MSe=summary(cloth.aov)[[1]][3,3]
```

$H_0 : T_1 = ... = T = p = 0$

vs

$H_1 : T_i \neq 0$ at least one.

From our ANOVA we see that our chemical has an f-value of 2.376 with a p-value = 0.121. Our p-value is too big so we must fail to reject our null hypothesis and conclude that there is no differences between the treatment/chemical types.

4.7

Consider the hardness testing experiment described in Section 4.1. Suppose that the experiment was conducted as described and that the following Rockwell C-scale data (coded by subtracting 40 units) obtained:

```
library(readxl)
cscale <- read_excel("cscale.xlsx")

cscaleLong=reshape(cscale,
                   varying = c("Coupon1", "Coupon2", "Coupon3",
                               "Coupon4"),
                   v.names = "Yield", timevar = "Coupon",
                   direction = "long")
```

(a) Analyze the data from this experiment.

```
cscale.aov=aov(Yield ~ factor(Coupon) + factor(Tip), data = cscaleLong)
summary(cscale.aov)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)    
## factor(Coupon)  3  0.825 0.27500   30.94 4.52e-05 ***
## factor(Tip)     3  0.385 0.12833   14.44 0.000871 ***
## Residuals       9  0.080 0.00889                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
MSe=summary(cscale.aov)[[1]][3,3]
```

$H_0 : T_1 = ... = T = p = 0$

vs

$H_1 : T_i \neq 0$ at least one.

Based on our ANOVA we see that we have a f-value of 14.44 and a p-value of 0.000871. Our p-value is very small compare to the significance value so we must reject our null hypothesis and conclude that the treatment/tips are significantly different.

(b) Use the Fisher LSD method to make comparisons among the four tips to determine specifically which tips differ in mean hardness readings.

```
ybar.trt=as.vector(with(cscaleLong, tapply(Yield,Tip,function(x) mean(x)))) #bar(Y_i.) vector

#1-2
a=4;b=4;N=a*b;alpha=0.05
t0=(ybar.trt[1]-ybar.trt[2])/sqrt(2*MSe/b)
cri=qt(alpha/2,(a-1)*(b-1),lower.tail=F)
pvalue=2*pt(abs(t0),(a-1)*(b-1),lower.tail=F); pvalue
```

```
## [1] 0.7163449
```

```
#1-3
a=4;b=4;N=a*b;alpha=0.05
t0=(ybar.trt[1]-ybar.trt[3])/sqrt(2*MSe/b)
cri=qt(alpha/2,(a-1)*(b-1),lower.tail=F)
pvalue=2*pt(abs(t0),(a-1)*(b-1),lower.tail=F); pvalue
```

```
## [1] 0.09354966
```

```r
#1-4
a=4;b=4;N=a*b;alpha=0.05
t0=(ybar.trt[1]-ybar.trt[4])/sqrt(2*MSe/b)
cri=qt(alpha/2,(a-1)*(b-1),lower.tail=F)
pvalue=2*pt(abs(t0),(a-1)*(b-1),lower.tail=F); pvalue
```

```
## [1] 0.001488949
```

```r
#2-3
a=4;b=4;N=a*b;alpha=0.05
t0=(ybar.trt[2]-ybar.trt[3])/sqrt(2*MSe/b)
cri=qt(alpha/2,(a-1)*(b-1),lower.tail=F)
pvalue=2*pt(abs(t0),(a-1)*(b-1),lower.tail=F); pvalue
```

```
## [1] 0.05100326
```

```r
#2-4
a=4;b=4;N=a*b;alpha=0.05
t0=(ybar.trt[2]-ybar.trt[4])/sqrt(2*MSe/b)
cri=qt(alpha/2,(a-1)*(b-1),lower.tail=F)
pvalue=2*pt(abs(t0),(a-1)*(b-1),lower.tail=F); pvalue
```

```
## [1] 0.002578639
```

```r
#3-4
a=4;b=4;N=a*b;alpha=0.05
t0=(ybar.trt[3]-ybar.trt[4])/sqrt(2*MSe/b)
cri=qt(alpha/2,(a-1)*(b-1),lower.tail=F)
pvalue=2*pt(abs(t0),(a-1)*(b-1),lower.tail=F); pvalue
```

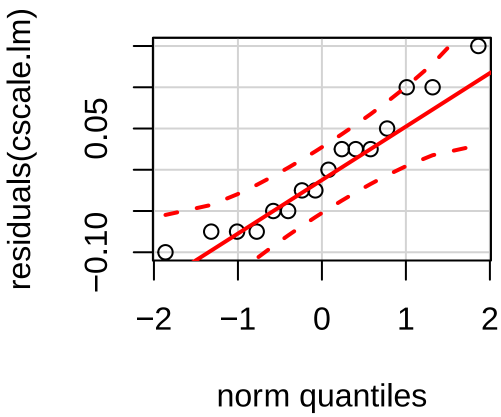
```
## [1] 0.0001290132
```

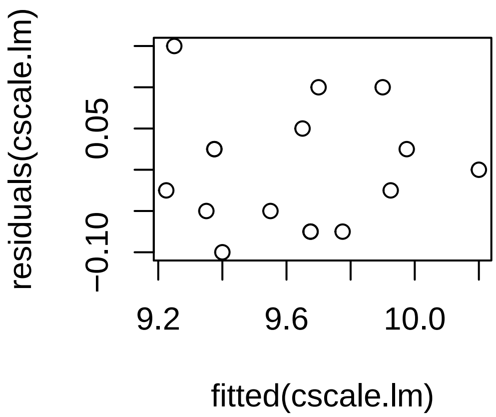
While using Fisher LSD we output a pvalue for the compariosns of $\mu_1 = \mu_2$, $\mu_1 = \mu_3$, $\mu_1 = \mu_4$, $\mu_2 = \mu_3$, $\mu_2 = \mu_4$, and $\mu_3 = \mu_4$ and the results are presented above respectively. We see that the comparisons of $\mu_1 = \mu_4$, $\mu_2 = \mu_4$, and $\mu_3 = \mu_4$ are different in tips of mean hardness.

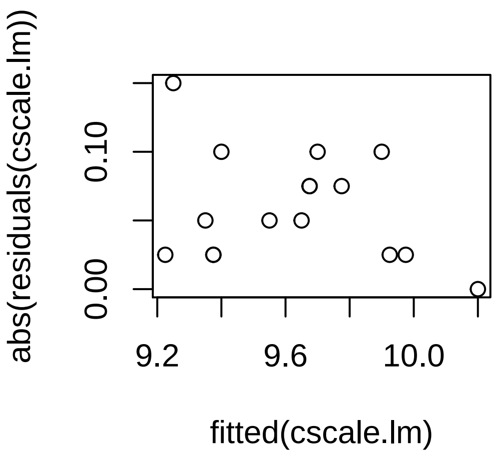
(c) Analyze the residuals from this experiment.

```
##
## Call:
## lm(formula = Yield ~ factor(Coupon) + factor(Tip), data = cscaleLong)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10000 -0.05625 -0.01250  0.03125  0.15000
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      9.35000    0.06236 149.934  < 2e-16 ***
## factor(Coupon)2  0.02500    0.06667   0.375 0.716345
## factor(Coupon)3  0.32500    0.06667   4.875 0.000877 ***
## factor(Coupon)4  0.55000    0.06667   8.250 1.73e-05 ***
## factor(Tip)2     0.02500    0.06667   0.375 0.716345
## factor(Tip)3    -0.12500    0.06667  -1.875 0.093550 .
## factor(Tip)4     0.30000    0.06667   4.500 0.001489 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.09428 on 9 degrees of freedom
## Multiple R-squared:  0.938,  Adjusted R-squared:  0.8966
## F-statistic: 22.69 on 6 and 9 DF,  p-value: 5.933e-05
```







After observing the qqplot and residual plot we see that normality is fine and there is no pattern but a structureless image on our plot. We can conclude our model is good.

4.13

Consider the ratio control algorithm experiment described in Section 3.8. The experiment was actually conducted as a randomized block design, where six time periods were selected as the blocks, and all four ratio control algorithms were tested in each time period. The average cell voltage and the standard deviation of voltage (shown in parentheses) for each cell are as follows:

```r
library(readxl)
rbdvolt <- read_excel("rbdvolt.xlsx")
rbdvoltLong=reshape(rbdvolt,
                    varying = c("Time1", "Time2", "Time3",
                                "Time4", "Time5", "Time6"),
                    v.names = "Yield", timevar = "Time",
                    direction = "long")

rbdsd <- read_excel("rbdsd.xlsx")
rbdsdLong=reshape(rbdsd,
                    varying = c("Time1", "Time2", "Time3",
                                "Time4", "Time5", "Time6"),
                    v.names = "Yield", timevar = "Time",
                    direction = "long")
```

(a) Analyze the average cell voltage data. (Use $\alpha = 0.05$.) Does the choice of ratio control algorithm affect the average cell voltage?

```r
rbdvolt.aov=aov(Yield ~ factor(Ratio.Control) + factor(Time), data = rbdvoltLong)
summary(rbdvolt.aov)
```

```
##                       Df  Sum Sq  Mean Sq F value Pr(>F)
## factor(Ratio.Control)  3 0.00275 0.000915   0.190  0.901
## factor(Time)           5 0.01744 0.003487   0.725  0.615
## Residuals             15 0.07218 0.004812
```

$H_0 : T_1 = ... = T = p = 0$

vs

$H_1 : T_i \neq 0$ at least one.

We look at our ANOVA and pay close attention to our results in terms of ratio control. We have a f-value of 0.190 and a p-value too big of 0.901. We fail to reject our null hypothesis and conclude that ratio control does not affect voltage.

(b) Perform an appropriate analysis on the standard deviation of voltage. (Recall that this is called "pot noise.") Does the choice of ratio control algorithm affect the pot noise?

```r
rbdsd.aov=aov(Yield ~ factor(Ratio.Control) + factor(Time), data = rbdsdLong)
summary(rbdsd.aov)
```

```
##                       Df   Sum Sq  Mean Sq F value   Pr(>F)
## factor(Ratio.Control)  3 0.026013 0.008671  50.756 4.34e-08 ***
## factor(Time)           5 0.002721 0.000544   3.185   0.0371 *
## Residuals             15 0.002563 0.000171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

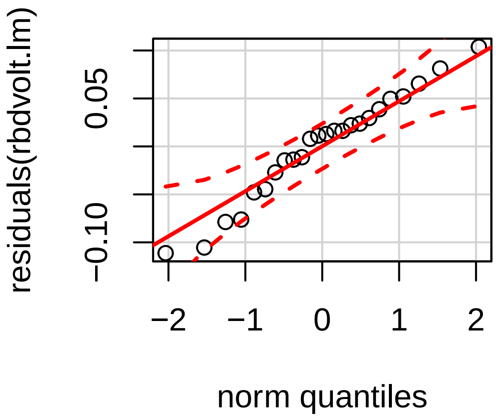
$H_0 : T_1 = ... = T = p = 0$

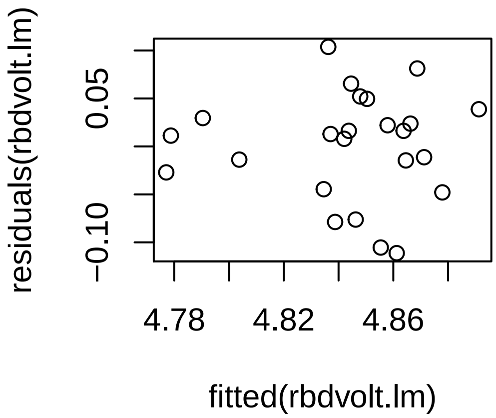
vs

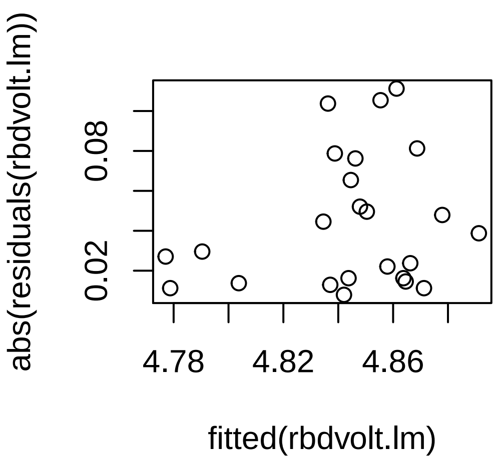
$H_1 : T_i \neq 0$ at least one.

From the ANOVA, observing the standard deviation of Voltage we see that we have f-value of 50.756 and a very small p-value of 4.34e-08 for ratio control. We reject our null hypothesis and conclude that ratio control does affect standard deviation or “pot noise”.

(c) Conduct any residual analyses that seem appropriate







By performing a qqplot and residual plots we see that normality is up to check and our residual plot is a structureless pattern which means that our model is good.

(d) Which ratio control algorithm would you select if your objective is to reduce both the average cell voltage and the pot noise?

From the first algorithm we were unsuccesful to reduce the voltage, on the other hand on algorithm two we manage to reduce the SD or Pot Noise successfully. It is best to choose algorithm 2.

4.18

Suppose that the observation for chemical type 2 and bolt 3 is missing in Problem 4.3. Analyze the problem by estimating the missing value. Perform the exact analysis and compare the results.

```
clothMissing=clothLong
clothMissing["2.3", "Yield"]=NA
summary(aov(Yield ~ factor(Chemical) + factor(Bolt), clothMissing))
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## factor(Chemical)  3  18.03    6.01   3.038   0.0747 .
## factor(Bolt)      4 140.84   35.21  17.797 9.06e-05 ***
## Residuals        11  21.76    1.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```

```
x=predict(aov(Yield ~ factor(Chemical) + factor(Bolt), clothMissing), newdata = clothLong["2.3", ]); x
```

```
##     1
## 75.25
```

```
#72.25
clothImpute=clothMissing
clothImpute["2.3", "Yield"]=75.25

a=aov(Yield ~ factor(Chemical) + factor(Bolt), clothImpute)
summary(a)
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## factor(Chemical)  3  12.78    4.26    2.35    0.124
## factor(Bolt)      4 158.89   39.72   21.90 1.92e-05 ***
## Residuals        12  21.76    1.81
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
a$df.residual=a$df.residual - 1 #important: correction of df!
summary(a)
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## factor(Chemical)  3  12.78    4.26   2.154    0.151
## factor(Bolt)      4 158.89   39.72  20.078 5.14e-05 ***
## Residuals        11  21.76    1.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Using R we are able to set the missing value as x and solve it, which results to 75.25. In which this value gives us the minimum error. As we can see from our new anova we see that all p-values and f-values are to check and can confirm this model is a good model.

4.22

The effect of five different ingredients (A, B, C, D, E) on the reaction time of a chemical process is being studied. Each batch of new material is only large enough to permit five runs to be made. Furthermore, each run requires approximately one and a half hours, so only five runs can be made in one day. The experimenter decides to run the experiment as a Latin square so that day and batch effects may be systematically controlled. She obtains the data that follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

```
library(readxl)
chemical <- read_excel("chemical.xlsx")

matrix(chemical$Ingrediants, 5,5)
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,] "a"  "b"  "d"  "c"  "e"
## [2,] "c"  "e"  "a"  "d"  "b"
## [3,] "b"  "a"  "c"  "e"  "d"
## [4,] "d"  "c"  "e"  "b"  "a"
## [5,] "e"  "d"  "b"  "a"  "c"
```

```
matrix(chemical$Reaction.Time, 5,5)
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    8    7    1    7    3
## [2,]   11    2    7    3    8
## [3,]    4    9   10    1    5
## [4,]    6    8    6    6   10
## [5,]    4    2    3    8    8
```

```
summary(aov(Reaction.Time ~ factor(Batch) + factor(Day) + Ingrediants, chemical))
```

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## factor(Batch)  4  12.24    3.06   0.979 0.455014
## factor(Day)    4  15.44    3.86   1.235 0.347618
## Ingrediants    4 141.44   35.36  11.309 0.000488 ***
## Residuals     12  37.52    3.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$H_0 : T_1 = ... = T = p = 0$

vs

$H_1 : T_i \neq 0$ at least one.

Based on our ANOVA we see that our treatment has a f-value of 11.309 with a p-value of 0.000488. Our p-value is very small and we reject our null hypothesis. We can state that chemical process and time of day have an affect on our ingrediants (A,B,C,D,E).